

PATENT ABSTRACTS OF JAPAN

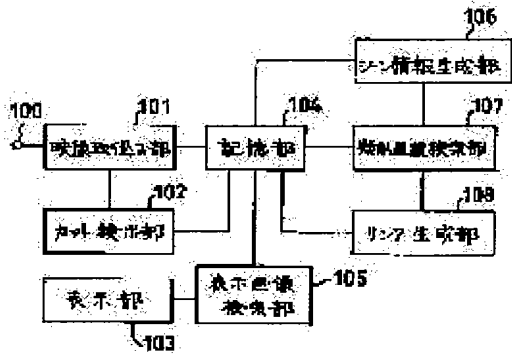
(11)Publication number : 09-093588
(43)Date of publication of application : 04.04.1997

51)Int.Cl. H04N 7/32
H04N 5/76
H04N 5/92

21)Application number : 07-250405 (71)Applicant : TOSHIBA CORP
22)Date of filing : 28.09.1995 (72)Inventor : KANEKO TOSHIMITSU
AOKI HISASHI

54) MOVING IMAGE PROCESSING METHOD

57)Abstract:
PROBLEM TO BE SOLVED: To improve the visual recognition by correcting a motion vector obtained at moving image compression into a vector representing a motion between consecutive frames and detecting an image of a frame satisfying a prescribed condition for a degree of dispersion in a frame generated next as a cut of a moving image.
SOLUTION: Moving image data are stored in a storage section 104 and a cut detection section 104 detects a cut already and information to be divided into a shot is provided. The user selects how a moving image is to be divided into scenes, by using what image. When a frame is selected and divided into scenes, a similar image retrieval section 107 detects them to be plural similar images. A scene information generating section 106 divides plural shots into scenes by using a frame discriminated to be a similar image as a border. A scene information generating section 106 divides plural shots into scenes, bordering the frame decided to be a similar image. When the same scene is discriminated, it is stored in a storage section 104 and plural frames being similar images are defined by a link generating section 108.



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平9-93588

(43)公開日 平成9年(1997)4月4日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所	
H 0 4 N	7/32		H 0 4 N	7/137	Z
	5/76			5/76	B
	5/92			5/92	H

審査請求 未請求 請求項の数10 O L (全 23 頁)

(21)出願番号 特願平7-250405

(22)出願日 平成7年(1995)9月28日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 金子 敏充

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(72)発明者 青木 恒

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

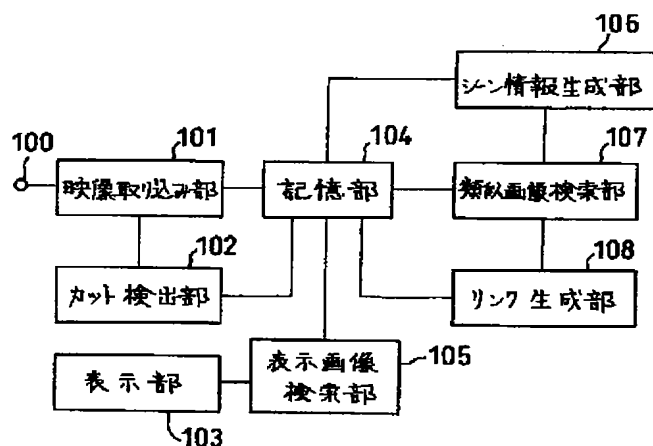
(74)代理人 弁理士 外川 英明

(54)【発明の名称】 動画像処理方法

(57)【要約】

【課題】 本発明は、動画像を意味構成に基づいて時分割する際の計算負担を軽減にし、そのようにして時分割された動画像部分を階層的に再構成して表示することにより、利用者が長時間の動画像の中から目的の場面により容易に到達できるようにすることを課題とする。

【解決手段】 動画像圧縮時に得られる動きベクトルデータの挙動を解析することで画像内容の変化点を検出し、動画像を時分割する。さらに時分割した各区間の画像内容の類似度を計算し、複数の分割区間の関連を階層的に定義する。利用者は所望の階層を上下させて動画像内容一覧の情報量を制御したり、画像中から抽出された登場物体を指示するなどして、目的の場面に到達する。



【特許請求の範囲】

【請求項1】動き補償を用いた動画像圧縮時に得られる動きベクトルを、連続するフレーム間の動きを表す動きベクトルに修正し、

修正された前記動きベクトルのフレーム内のばらつきの度合いを生成し、

生成された前記ばらつきの度合いが所定の条件を満たすフレームの間を、動画像のカットとして検出することを特徴とする動画像処理方法。

【請求項2】フレーム群に分割されている動画像の任意のフレーム群と他のフレーム群の間の類似度を求め、求めた前記類似度が所定の条件を満たすときに前記フレーム群をリンク付け、リンク付けした前記フレーム群の間を分割することを特徴とする動画像処理方法。

【請求項3】フレーム群に分割されている動画像の任意のフレーム群と他のフレーム群の間の類似度を求め、求めた前記類似度が所定の条件を満たすときに前記フレーム群をリンク付け、リンク付けした前記フレーム群を同一の集合に属するものとすることを特徴とする動画像処理方法。

【請求項4】フレーム群に分割されている動画像から文字やシンボルを認識し、認識した前記文字やシンボルのうち、関連した文字やシンボルを含むフレーム群をリンク付け、リンク付けした前記フレーム群の間を分割することを特徴とする動画像処理方法。

【請求項5】フレーム群に分割されている動画像から文字やシンボルを認識し、認識した前記文字やシンボルのうち、関連した文字やシンボルを含むフレーム群をリンク付け、リンク付けした前記フレーム群を同一の集合に属するものとすることを特徴とする動画像処理方法。

【請求項6】フレーム群に分割されている動画像の任意のフレームと他のフレーム群の間の類似度を求め、求めた前記類似度が所定の条件を満たすフレームを前記フレーム群の代表フレームとすることを特徴とする動画像処理方法。

【請求項7】フレーム群に分割されている動画像の各フレーム群中の文字またはシンボルを含んでいるフレームを代表フレームとすることを特徴とする動画像処理方法。

【請求項8】フレーム群に分割されてる動画像の区切られた複数の時間区間を一連として表示し、一連として表示された前記動画像の時間区間の区切りの位置を、時刻に対応する位置に表示することを特徴とする動画像処理方法。

【請求項9】動画像の視聴の進行に応じて、一連として表示された前記動画像の表示が変化することを特徴とする動画像処理方法。

【請求項10】フレーム群に分割されてる動画像の区切られた複数の時間区間を一連として表示し、一連として表示された前記動画像の特徴的な対象物の領域を認識して抽出し、

抽出した前記対象物の領域を、抽出を行った動画像の時刻に対応した位置に隣接させて表示することを特徴とする動画像処理方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、動画像の圧縮符号化時に映像カットを自動的に検出し、動画像を分割・階層化した情報を保持し、表示の際に利用する動画像処理方法に関する。

【0002】

【従来の技術】動画像をカットごとに区切っておくことにより、動画に対する編集や検索を容易にすることが可能になる。例えば、動画の早見を行う際に、カットの直後のフレームのみを選択して表示したり、表示を行うか否かをカットを区切りとして定めておいたりすることができる。また、動画に対してキーワードを設定しておく、動画の中の一部だけを見たい場合には結局動画全体を見なければならなくなってしまうが、カットにキーワードを付与しておけば、簡単に動画の中の所望の部分にすばやくアクセスすることができる。編集時には動画像の一部分の削除や追加はカットを区切りとして行うことが多いので、動画が初めからカットごとに区切ってあれば作業が容易に行える。

【0003】動画像のカットは動画像の意味的な構造化にも利用することができる。カットから次のカットまでの映像は一連の連続した内容を持っており、動画像を構造化するにはカットから次のカットまでの映像が動画像の最小単位となる。動画像の構造化は、動画像の編集や検索において大変に有意義なことである。例えば、全ての映像を見ることなく目的に関連した場面のみをすばやく表示させるために利用できる。また、動画の構造を視聴者にあらかじめ提示しておくことにより、動画の内容理解を容易たらしめることもできる。

【0004】しかしながら、動画をカットごとに区切るという作業は、作業を行うものにとって大変な負担となっています。カットを見つける作業を行うためには、作業者は常に動画を注視していなければならない。しかも、映像のカットを発見しても正確なカット点を見つけるためには動画の逆再生、低速再生、一時停止などの操作を駆使しなければならない。このような作業を行うには、動画像の再生時間の何十倍もの作業時間を要することになってしまう。従って、動画のカットごとの区分は自動化する必要がある。

【0005】動画の自動カット検出手法はこれまでもいくつか考案されている。例えば、連続するフレーム間の対応する画素の輝度差を求め、フレーム全体の総和を

とり、この総和としきい値との比較によりカットを検出する手法がある。しかしながら、このような手法は動画中の背景やオブジェクトといった内容をあまり考慮していない。従って、どうしても誤検出や未検出が生じてしまう。

【0006】一方、動画中の背景やオブジェクトをはっきりと認識することができれば正確にカットの検出ができるのであるが、動画の自動内容理解をしようとするとはく大な計算時間がかかり、しかも不十分な結果しか得ることができない。

【0007】また、従来、動画画は連続するフレーム全体で一つの単位をなしており、単に一つの動画画を選択して再生するのが一般的であった。したがって、動画画中の所望の場面を選択して再生したい場合には、まず、一本の動画画を選択して再生し、時間的に未来方向、過去方向に早送りしながら所望の場面まで到達していた。すなわち、動画画のはじめから、または、たまたま選択された途中の場面から動画画の時間軸をたどって所望の場面まで到達していた。また、所望の場面が動画画のどのあたりにあったかをあらかじめ記憶している場合には、その記憶を頼りにして再生開始位置を指定し、到達するまでの時間を短くするようにしていた。

【0008】以上は以前見たことのある動画画から所望の場面を探す場合の手続きであったが、一度も見たことのない動画画から利用者が興味のあるような場面を選択するためには、動画画のはじめから再生するしかない。動画を見る時間を短縮するためには、早送り操作を織り交ぜながら動画画を早見したり、時間的に未来方向または過去方向に再生開始場所を適当に飛ばして再生し、興味のあるような場面が再生されたら通常の再生速度で再生し、興味がなさそうな場面であれば再び再生開始場所を飛ばして次の場面の再生へと移っていた。

【0009】もう少し効率よく所望の場面を探すための手法として、動画画をいくつかの場面に分割しておき、それぞれの場面を代表するフレームを一覧表示するものがある。利用者は表示された代表フレームの中から所望の場面に相当する代表フレームを選択し、その場面から動画画を再生する。また、まだ見たことのない動画画から利用者の興味のある場面を選択してみる場合には、代表フレームの一覧を見ることにより動画画の内容をざっと見ることができると、比較的容易に興味のある場面を予測して、その場面からの再生を行うことができる。

【0010】昨今の家庭用ビデオカメラの普及、放送メディアの多様化、映像ソフトのレンタルなどにより、一般家庭においても大量の映像情報を享受できるような環境になりつつある。さらに、現在各方面で研究開発や実験が行われているビデオ・オン・デマンド（VOD）などが実現すれば、扱うことのできる映像の量はますます増加することは明らかである。

【0011】一方、家庭向けのパーソナルコンピュータ

の高機能化もめざましく、従来、文字と単純な図形しか扱えなかったのに対し、最近では音声や映像、あるいは動画画を表示したり編集したりする機能をもったものも比較的安価に入手できるようになっている。また、動画画はコンピュータの記憶領域を大量に占有するが、CD-ROMの大量生産やハードディスクの低価格化、あるいは記録可能な光ディスクの実現などにより、家庭でも大量記憶装置を接続するユーザーが増えている。

【0012】このような環境のもと、パーソナルコンピュータ上、あるいはパーソナルコンピュータ程度の能力を持った専用ビデオ装置で大量の動画画を処理するようなシステムが登場すると考えられる。

【0013】しかし動画画はそのままでは文書や静止画と異なり、一覧性がない。つまり、いちどきに動画画の全体のストーリーを把握することができない。すばやく内容を把握しようとするとき、私たちはビデオを（画面を提示させながら）早送りしたり、ある時間間隔でとびとびに見たりするが、これは全体を把握するには時間を要する。

【0014】そこで一連の動画画から、特定の瞬間の画像を抽出し、それを小画面にして複数枚並べた一覧表

（図2参照）で全体を把握するという方法が考えられる（参考文献：「認識技術を応用した対話型映像編集方式の提案」電子情報通信学会論文誌 D-I I Vol. J75-D-I I No. 2 pp. 216-225 1992年2月）。

【0015】この「特定の瞬間」とは、5秒、30秒、など、画像内容に関知しないものでもよく、画像の変化点を用いてもよい。画像の変化点には従来「カット」が用いられてきた。カットは撮影時のカメラ動作が一時停止した時刻であり、ある瞬間前後の画像を比較して、差が大きい時にカットが発生したとみなす方法などで自動検出する技術も発表されている。

【0016】しかし、映画などの動画画の場合、カットは平均5秒程度ごとに発生するといわれている。すなわち、自動検出が正確に行われたとすると2時間の映画では1440回となる。と、なると画面上に1440の小画面が表示されることになり、操作性が著しく低下することは言うを待たない。また、カット時の画像を一覧表示に用いると、次のカットまでの間に登場して消滅したオブジェクト（対象物、被写体）は一覧に現れないため、こういったオブジェクトを目的として映像の検索をする事が困難である。

【0017】

【発明が解決しようとする課題】従来の映像カット自動検出技術では、前述したような誤検出や未検出が多く発生してしまう。誤検出や未検出は最終的には作業によって修正されねばならず、結局作業者が動画全般を検査しなければならなくなり、カット検出を自動化した意味が半減してしまうという問題がある。また、完全な動画

の自動内容理解を行うことは技術的に不可能である。不十分な自動内容理解であっても、そのための計算にばく大な時間がかかってしまい、実際には利用できなくなってしまう。本発明の目的は、以上の問題点を解決し、計算をあまり複雑にすることなく、自動に、しかも正確に映像カットの検出を行うことである。

【0018】従来のような動画像の時間軸をたどりながらの検索では、動画像の時間軸上の一点しか利用者が見ることができないために所望の場面にたどり着くまでの時間が非常に長くなり、効率的ではない。特に、長い動画像の場合には顕著である。利用者の記憶を頼りにして動画像の再生開始位置を指定し、所望の場面を探す場合にも、利用者の記憶は曖昧で信頼の置けないものであり、記憶違いの場合にはかえって検索の効率を悪くしてしまうことさえある。また、所有している動画像全てに対してどの場面が動画像のどのあたりに存在していたかを記憶しておくのは利用者に多大な付加をかけることになるし、多くの動画像にたいしては全て記憶すること事態が不可能である。

【0019】一方、動画像の中から代表フレームを選び、これを一覧表示して所望の場面を探す手法においては、動画像の時間軸上の複数の点の映像を同時に見ることができると、検索の効率は改善されている。しかしながら、代表フレームを利用者自身が選択してあらかじめ登録しておくのは、利用者に負担を強いることになる。この負担をなくすために、自動的に動画像の不連続点（ショットの切れ目）を検出し、それぞれのショットの中から代表フレームを選ぶ（例えばショットのはじめのフレームを選ぶ）手法があるが、編集された映像ではショットは一般に5秒程度と短いため、表示される代表フレーム数が多くなってしまうという欠点がある。

【0020】画面に多くの静止画が一覧表示されると、かえって動画像の内容がわかりづらくなり、利用者が困惑してしまう場合がある。また、ときには画面上に表示しきれなくなり、複数回に分けて表示しなければならなかったり、一度に表示しようとするために一つの代表フレームの表示がきわめて小さくなってしまったりする。このような場合には、所望の場面の検索効率の改善度が低下してしまうという問題がある。

【0021】従来の技術では単にカットを検出し、それを小画面として列記するだけであったために、列記の数が増えることがある、ショット内の画像内容が不明確になることがある、などの問題点があった。本発明はカットによって切り分けられた画像の一覧表示の際に、複数のショットを統合して小画面の列挙を減らすことにより視認性を高め、また、統合した場合でもショット内の画像を効率よく把握できるようにすることを目的とする。

【0022】

【課題を解決するための手段】本願発明は、動き補償を用いた動画像圧縮時に得られる動きベクトルを、連続す

るフレーム間の動きを表す動きベクトルに修正し、修正された前記動きベクトルのフレーム内のばらつきの度合いを生成し、生成された前記ばらつきの度合いが所定の条件を満たすフレームの間を、動画像のカットとして検出することを特徴とする。

【0023】また、フレーム群に分割されている動画像の任意のフレーム群と他のフレーム群の間の類似度を求め、求めた前記類似度が所定の条件を満たすときに前記フレーム群をリンク付け、リンク付けした前記フレーム群の間を分割することを特徴とする。

【0024】さらに、フレーム群に分割されてる動画像の区切られた複数の時間区間を一連として表示し、一連として表示された前記動画像の時間区間の区切りの位置を、時刻に対応する位置に表示することを特徴とする。

【0025】これにより、本発明の動画像処理装置における自動映像カット検出技術は、動画における背景やオブジェクトの動きを示す動きベクトルの空間的、または時間的連続性をもとに映像カットの検出を行うので、動画中の意味情報を生かした正確な映像カット検出が自動的に行える。また、動き補償を用いた動画像圧縮時に算出される動きベクトル情報をもとに検出を行うため、カット検出のためのばく大な計算を必要とせず、比較的少ない計算で実現することができる。

【0026】また、動画像が細かく分割されているときに、これら分割された部分のいくつかをまとめてより大きく分割された動画像を構成し、それぞれの部分について一つの代表フレーム、もしくはそれに関する情報を表示することができるので、表示される情報が多くなりすぎることがなく、利用者に負担をかけずに検索が行えるようになる。

【0027】さらに、利用者がこれら分割を自ら行ったり、代表フレームもしくはそれに関わる情報を一つ一つ選択していくことが不要であるため、利用者の手を煩わせることがない。

【0028】その上、複数のショットを一括して表示するため、カットの多い画像を一覧表示した際にも、一覧画像の数が増えず、視認性の向上が期待できる。また、このようにして一括しても、ショット内の情報を奥行き、あるいは奥行きに隣接する位置に表示するため、ショットの内容を円滑かつ迅速に把握することができる。と期待できる。

【0029】

【発明の実施の形態】

（第一の実施例1）以下、本発明の実施例を図に基づいて説明する。図1は本発明を実現するためのブロック構成の一実施例である。

【0030】図1において映像入力端子1から、映像信号は入力される。DCT部2は、入力された映像信号と動き補償部6により予測された映像信号との差分映像信号をある決められたブロック単位に分割し、離散コサイ

ン変換(DCT)を行う。離散コサイン変換により差分映像信号は少ない低域係数に集中するため、画像の空間的情報量が削減される。

【0031】量子化部3は、離散コサイン変換された差分映像信号を量子化する。量子化により離散コサイン変換された差分映像信号全体が少ない数の集合で表現されるため、符号量が減らされる。

【0032】可変長符号化部4は、量子化された差分映像信号を異なる長さの信号の集合で表現することにより、統計的に平均情報量の削減を行う。圧縮映像出力端子11は、圧縮された映像信号が出力される。

【0033】逆量子化部5は、量子化部3と逆の信号変換を行う。DCT部2の出力信号と逆量子化部5の出力信号とは、量子化誤差の分だけ異なっている。逆DCT部8は、DCT部2と逆の信号変換を行い、離散コサイン変換された差分映像信号を元の信号に復元する。ただし、復元された信号は誤差を含んでいる。

【0034】フレームメモリ7は、逆DCT部8により復元された差分映像信号と動き補償部6の出力映像信号とを加算した映像信号をフレーム単位で数フレーム分記憶しておくメモリである。

【0035】動きベクトル検出部9は、フレームメモリ7に記憶されている映像信号と新たに映像入力端子1から入力される映像信号とを比較し、二つの映像間の動きベクトルを検出する。すなわち、新たに入力された映像のフレームを所定のブロックに分割し、それぞれのブロックが、もう一方のフレームメモリに記憶されたフレーム映像中のどの部分に移動したかを検出し、その移動ベクトルをブロックの動きベクトルとする。

【0036】動きベクトルの検出方法としては、二つのブロック間の類似度を定義し、最も大きい類似度を与える移動ベクトルをブロックの動きベクトルとしている。これはフレーム単位ではなく、フィールド単位で行われてもよい。動きベクトルは1フレームの映像につき、フレーム中のブロックの数だけ検出される。

【0037】動き補償部6は、フレームメモリに記憶された映像を動きベクトル検出部9において検出された動きベクトル分だけブロックごと移動させて映像を構成する。このようにして構成された映像は予測映像として出力される。

【0038】映像カット検出部10は、動きベクトル検出部から出力される動きベクトルからフレーム間の変化量を算出して映像のカット情報を出力する。映像カット情報の出力端子12である。以下では、映像カット検出部10の構成および動作をより詳細に説明する。

【0039】図2は映像カット検出部10のブロック図である。動きベクトル修正部21は、動きベクトル検出部9から出力された動きベクトルを修正し、連続するフレーム間の動きベクトルを算出する。

【0040】映像変化量算出部22は、動きベクトル修

正部21から出力された連続する2フレーム間の動きベクトルから、これら2フレーム間の映像の変化量を算出する。

【0041】映像変化量メモリ24は、映像変化量算出部22の出力である映像変化量を数フレーム分記憶しておくためのメモリである。しきい値算出部25は、映像変化量メモリに記憶された過去数フレーム分の映像変化量からしきい値を算出し、比較部23へ出力する。

【0042】比較部23は、映像変化量算出部22から出力された映像変化量としきい値算出部25から出力されたしきい値との比較を行い、フレーム間にカットがあったか否かを判定し、映像カット情報として出力する。

【0043】以下、さらに詳細に映像カット検出部の動作を説明する。まず、映像変化量算出部の動作を説明する。動きベクトル修正部21から出力された動きベクトルは連続する2つのフレーム間の動きベクトルに修正されている。これを修正動きベクトルと呼ぶことにする。修正動きベクトルは、1つのフレームあたり、動きベクトルを求める際にフレームを分割したブロックの数だけ存在している。これら複数存在する修正動きベクトルを、さらに適当な大きさのブロックに分割する。図3は3×3の動きベクトルのブロックに分割した例を示しており、31は動きベクトルを求める際に分割された1ブロック、32は1フレーム中の動きベクトルを分割したブロックである。以下、1フレーム中の動きベクトルを分割したブロックを動きベクトルブロックと呼ぶことにする。

【0044】フレームtの動きベクトルブロック $B_t(i)$ 、 $i=1, 2, \dots, NB$ とする。NBは1フレーム中の動きベクトルブロック数部ある。それぞれの動きベクトルブロックの映像変化量 $\Delta t(i)$ を

【0045】

【数1】

$$\Delta t(i) = \min (\Delta^+ t(i), \Delta^- t(i))$$

として求める。ただし、 $\Delta^+ t(i)$ しか存在しない場合には $\Delta t(i) = \Delta^+ t(i)$ とし、 $\Delta^- t(i)$ しか存在しない場合には $\Delta t(i) = \Delta^- t(i)$ として求めるものとし、 $\Delta^+ t(i)$ 、 $\Delta^- t(i)$ は

【0046】

【数2】

$$\Delta^+ t(i) = \sum |v^+_{t,i} - v^+_{t(j)}|^2$$

$$v^+_{t(j)} \in B_t(i)$$

$$\Delta^- t(i) = \sum |v^-_{t,i} - v^-_{t(j)}|^2$$

$$v^-_{t(j)} \in B_t(i)$$

のように求めるものとする。ここで、 $v^+_{t(j)}$ 、 $j=1, 2, \dots, Nn$ はフレームtの順方向の修正動き

ベクトル、 $v-t(j)$ 、 $j=1, 2, \dots, N_n$ は n は動きベクトルブロック内の逆方向の修正動きベクトル数であり、後に説明する。 $v+t(j)$ または $v-t(j)$ に常に定義されているとは限らず、定義されていない場合には $\Delta+t(i)$ または $\Delta-t(i)$ は存在しないものとする。また、 $v+t, i$ 、 $v-t, i$ は $Bt(i)$ に含まれる修正動きベクトルの平均ベクトルで、

【0047】

【数3】

$$v^{+}t, i = \sum_{v^{+}t(j) \in Bt(i)} v^{+}t(j) / n$$

$$v^{-}t, i = \sum_{v^{-}t(j) \in Bt(i)} v^{-}t(j) / n$$

により求められる。動きベクトルブロックの変化量からフレーム単位の映像変化量 Δt を

【0048】

【数4】

$$\Delta t = \sum \Delta t(i)$$

として求める。

【0049】動きベクトルブロックの変化量 Δt は、動きベクトルブロック内に含まれる修正動きベクトルのばらつきを示す量である。図4はフレーム t および一つ前のフレーム $t-1$ が連続した映像の場合の修正動きベクトル、図5はフレーム $t-1$ とフレーム t との間に映像カットがあった場合の修正動きベクトルの例を示している。映像カットがなく連続した映像の場合には、映像内の背景やオブジェクトはほぼ同じ修正動きベクトルを有するため、動きベクトルブロックの変化量 $\Delta t(i)$ は小さな値を示す。一方、フレーム $t-1$ とフレーム t との間に映像カットがあった場合には、修正動きベクトルはもはや意味を持たず、相関のない映像間の動きベクトルを無理やり求めるために動きベクトルブロックの変化量 $\Delta t(i)$ は大きな値を示す。このような理由により、動きベクトルブロックごとの変化量 $\Delta t(i)$ の総和 Δt は、フレーム $t-1$ とフレーム t の間に映像カットが存在する場合に大きな値、存在しない場合には小さな値を得るため、 Δt を利用して映像カットを検出することができる。

【0050】映像カットの判定は比較部23において行われる。比較部23は映像変化量算出部22から出力されるフレームごとの映像変化量としきい値算出部25から出力されるしきい値との比較を行い、図9に示すようにしきい値の方が大きい場合に映像カット信号を出力する。

【0051】しきい値算出部25はしきい値を算出するブロックであるが、しきい値は映像変化量メモリ24に記憶された所定フレーム分の映像変化量を参照して、様

々な方法で計算される。例えば、最も単純なしきい値算出方法はしきい値を定数とする方法である。このような場合には、映像変化量メモリ24は取り除くことができる。また別の例としては、映像変化量メモリ24に記憶された映像変化量の平均値を求め、平均値の定数倍の値をしきい値とする方法がある。

【0052】さらに別の例としては、映像変化量メモリ24に記憶された映像変化量の平均 μ と分散 σ^2 を求め、しきい値を $\mu + c\sigma^2$ に設定する。ここで c はあらかじめ決められた定数である。また別の方法としては、映像変化量メモリ24に記憶された映像変化量の最大値と最小値との差を R 、 c' を正の定数として、しきい値を $\mu + c'R$ とする方法などがある。

【0053】動きベクトル修正部21は先に述べたとおり、動きベクトル検出部9から出力される動きベクトルを修正し、連続するフレーム間の動きベクトルを算出するためのブロックである。図6は動き補償を使った画像圧縮におけるフレームと検出される動きベクトルの関数を示したものである。図中、双方向フレーム内符号化画面51、フレーム間予測符号化画面52、53はフレーム内符号化画面である。

【0054】このようなフレーム構成の場合には、動きベクトル検出部9は図6中の矢印のフレーム間に対してのみ、動きベクトルが求められる。従って、任意の連続するフレーム間の動きベクトルは直接得ることはできない。そこで、動きベクトル修正部21は以下のように連続するフレーム間の動きベクトルを2通りの方法で求める。

【0055】まず、フレーム t の順方向の動きベクトルを $u+t(i)$ 、 $i=1, 2, \dots, N_n$ とする。このとき、フレーム t の順方向の修正動きベクトル $v+t(j)$ は、フレーム t とフレーム $t-1$ の順方向の動きベクトルの差分ベクトルとして $v+t(j) = u+t(i) - u+t-1(i)$ を求める。図7は順方向の修正動きベクトルを図示したものであり、図中61はフレーム t およびフレーム $t-1$ 、62は順方向の動きベクトル $u+t-1(i)$ 、63は順方向の動きベクトル $u+t(i)$ 、64は順方向の動きベクトル $v+t(j)$ である。ただし、フレームの構成によっては順方向の動きベクトル $u+t$ や $u+t-1(i)$ が存在しないことがあるが、 $u+t-1(i)$ が存在しない場合には $u+t-1(i)$ をゼロベクトルとして取り扱い、また、 $u+t(i)$ が存在しない場合には $v+t(j)$ を定義しないものとする。フレーム t の逆方向の修正動きベクトルを求める場合も、順方向の修正動きベクトルを求める場合とほぼ同様である。フレーム t の逆方向の動きベクトル $u-t(i)$ 、 $i=1, 2, \dots, N_n$ とする。フレーム t の逆方向の修正動きベクトル $v-t(j)$ は、フレーム t とフレーム $t-1$ の逆方向の動きベクトルの差分ベクトル $v-t(j) = u-t-1(i) - u-t(i)$ として求める。

【0056】図8は逆方向の修正動きベクトルを図示したものであり、65は逆方向の動きベクトル $u-t$

(i)、66は逆方向の動きベクトル $u-t-1(i)$ 、67は逆方向の修正動きベクトル $v-t(j)$ である。 $u-t(i)$ が存在しない場合には $u-t(i)$ をゼロベクトルとして取り扱い、また、 $u-t-1(i)$ が存在しない場合には $v-t(j)$ を定義しないものとする。

【0057】このように順方向、逆方向の二つの修正動きベクトルを生成することにより、順方向、または逆方向の一方の動きベクトルが映像カットの影響を受けている場合でも、もう一方の動きベクトルから映像カットの影響を受けない修正動きベクトルを生成することができる。その結果、映像カットの誤検出を低減させることが可能となる。

【0058】比較部23から出力された映像カット信号は、映像信号と関連づけて記録される。映像再生の際に映像カット信号を検出することによりカットごとに映像の頭出しをしたり、また、キーワードと映像カットの生じたフレームとを関連づけて記録しておき、カットごとのキーワード検索を行うことなどに利用できる。

【0059】(第一の実施例2)次に、第一の実施例と構成は同じであるが、映像変化量算出部22における映像変化量の算出方法において、異なる算出方法を用いた実施例を説明する。

【0060】まず、映像変化量算出部22では、動きベクトル修正部21からの出力である修正動きベクトルからフレーム t の順方向の修正動きベクトルの差分ベクトル $w+t(j)$ 、 $j=1, 2, \dots, Nn$ 、および逆方向の修正動きベクトルの差分ベクトル $w-t(j)$ 、 $j=1, 2, \dots, Nn$ を以下のように求める。

【0061】

【数5】

$$w^+t(j) = v^+t(j) - v^+t-1(j)$$

$$w^-t(j) = v^-t(j) - v^-t(j)$$

第二の実施例におけるフレームの映像変化量は、 $w+t(j)$ および $w-t(j)$ を利用して算出される。例えば、フレーム t の映像変化量 δt を

【0062】

【数6】

$$\delta t = \min(\sum^{Nn} |w^+t(j)|, \sum |w^-t(j)|)$$

として算出する。 $v+t-1(j)$ が定義されていないときには $v+t-1(j)$ はゼロベクトルとして取り扱われ、 $v+t(j)$ が定義されていないときには $w+t(j)$ も同様に定義されないものとする。同様に、 $v-t+1(j)$ が定義されていないときには $v-t+1(j)$ はゼロベクトルとして扱われ、 $v-t(j)$ が定義されていないときには w

$-t(j)$ は定義されないものとする。

【0063】 δt は動きベクトルの時間的な変化が小さいときに小さな値を取り、時間的な変化が大きいときに大きな値を取る。すなわち、映像中の背景やオブジェクトの動きがなめらかか否かの度合いを示している。フレーム t とフレーム $t-1$ との間に映像カットが存在する場合には背景やオブジェクトの動きがなめらかでなくなり、その結果動きベクトルの時間的な変化が大きくなる。従って δt が大きな値を示すことになり、第一の実施例と同様の方法でしきい値算出部25において算出されたしきい値との比較を行うことにより、映像カットを検出することができる。

【0064】(第一の実施例3)本発明の別の実施例は、図10のブロック図のごとく構成されることにより実現される。1から9までのブロックは図1中のそれと全く同一のものである。81は映像カット検出部10とは別の構成からなる映像カット検出部である。第一、第二の実施例と異なる点は、映像信号が直接映像カット検出部81に入力される点である。以下で映像カット検出部81の動作を説明する。

【0065】図11は映像カット検出部81の構成を示すブロック図である。動きベクトル修正部21は図2におけるそれと同一のものである。第一の映像変化量算出部91は、第一の実施例または第二の実施例で説明した映像変化量算出部であり、第一の実施例における映像変化量 Δt または第二の実施例における映像変化量 δt を第一の映像変化量として出力する。

【0066】第二の映像変化量算出部92は、映像信号から第二の映像変化量 $d t$ を算出する。具体的な算出方法としては、例えば(「輝度情報を使った動画ブラウジング」、大辻・外村・大庭、電子情報通信学会技術報告書、IE90-103、1991)に記されているように連続するフレーム間の画素単位の輝度差の総和を映像変化量とする方法や、(「動画像データベースハンドリングに関する検討」、外村・安部、電子情報通信学会技術報告書、IE89-33、1989)に記されているように、フレームごとに輝度ヒストグラムを求め、連続するフレーム間で輝度ヒストグラムの差分の和を映像変化量とする方法や、さらに(「カラービデオ映像における自動索引付け方と物体探索法」、長坂・田中、情報処理学会論文誌、pp.543-550、1992)に記述されているように連続する二つのフレーム映像のRGB信号に対して統計的処理に用いられるカイ二乗統計量を求め、映像変化量とする方法などがある。第二の映像変化量算出部92は、これらの手法を使って求められた映像変化量を第二の映像変化量として出力する。

【0067】映像変化量メモリ93は第一の映像変化量および第二の映像変化量を定められたフレーム分記録しておくメモリである。しきい値算出部94は映像変化量メモリ93に記録された数フレーム分の第一、および第

二の映像変化量から第一および第二の映像変化量に対するしきい値を算出する。

【0068】しきい値算出の手法としては、第一の実施例において説明したように、しきい値を定数としておく方法や、映像変化量メモリ93に記憶された過去の映像変化量の平均値、分散、最大値と最小値との差分により算出するなどの方法を使うことができる。しきい値の算出方法は第一の映像変化量と第二の映像変化量に対して独立に行ってよい。

【0069】判定部95は第一の映像変化量算出部91から出力された第一の映像変化量としきい値算出部94から出力された第一の映像変化量に対するしきい値とから、第一の映像変化量による映像カット検出を行う。さらに第二の映像変化量算出部91から出力された第二の映像変化量としきい値算出部94から出力された第二の映像変化量に対するしきい値とから、第二の映像変化量による映像カット検出を行う。そして二つの検出結果から最終的な映像カットの判定を行う。この判定方法としては、例えば、二つの検出結果がともに映像カットであった場合にのみ最終的に映像カットであると判定する方法や、どちらかの検出結果が映像カットであれば最終判定を映像カットとする方法などがある。

【0070】(第一の実施例による効果)本発明の動画画像処理装置における自動映像カット検出技術は、動画における背景やオブジェクトの動きを示す動きベクトルの空間的、または時間的連続性をもとに映像カットの検出を行うので、動画中の意味情報を生かした正確な映像カット検出が、自動的に行うことを可能とする。また、動き補償を用いた動画画像圧縮時に算出される動きベクトル情報をもとに検出を行うため、カット検出のためのばく大な計算を必要とせず、比較的少ない計算で実現することを可能にする。

【0071】(第二の実施例1)図13は本発明の動画画像処理装置の第二の実施例1の構成を表した図である。各部ブロックの役割を簡単に説明しておく。図13において、101は映像取り込み部で、動画画像データ入力端子100から入力された動画画像データを記憶部104に記憶するのに都合の良いデータに変換する。また、データ圧縮もここで行われる。入力データがアナログ信号である場合にはアナログ/デジタル変換も施される。102はカット検出部で、映像不連続点(以下カットと呼ぶ)の検出を行う。カットとカットの間は一つのショットを構成している。カットの検出はすでに記憶部104に蓄積されている動画画像データに対して行う場合と、動画画像データ入力端子100に入力されたデータを取り込みながら行う場合とがある。カットの検出には、フレームごとの輝度の分布や色の分布などの変化、また、動画画像データ圧縮時に求められるパラメータ等が利用される。具体的には、「映像カット自動検出方式の検討」

(大辻、外村、テレビジョン学会技術報告、vol.1

6、No.43、pp.7-12、1992)で検討されている各種手法が適用できる。カット検出の結果は動画画像のカット情報として動画画像データとともに記憶部104に蓄積される。

【0072】類似画像検索部は、動画画像の中の1フレームをキー映像として保持し、他のフレームとの類似度を計算する。あらかじめ定められた値以上の類似度を持つフレームが検出されたときには、その旨を示す信号をリンク生成部108及びシーン情報生成部106に送る。リンク生成部108は類似画像検索部107からの信号を受け、類似度したフレーム間に関連のあることを示すリンクを生成する。このリンクはフレーム間に限らず、カット情報により分割されたショット単位のリンクも生成する事ができる。リンク情報は記憶部104に蓄積される。シーン情報生成部106は類似画像検索部107からの信号を受け、類似したフレームまたはショットを境にしてその間のショット群を同一のシーンであると定義し、その情報を動画画像データとともに記憶部104に蓄積する。

【0073】表示画像検索部105は利用者の指示によって、表示部103に表示すべき映像を記憶部104から検索し、表示させる。一つのシーンごとに代表フレームを一つ検索し、全てのシーンの代表フレームを表示させたり、利用者によって一つのシーンが選択された場合にはそのシーンに含まれるショットに関して、各ショットを代表するフレームを表示させたりすることができる。

【0074】以下、具体的な例を用いてさらに詳しく説明する。記憶部104には動画画像データが蓄積されており、この動画画像はすでにカット検出部102によりカットの検出がなされ、ショットに分割するための情報を持っている。利用者は、動画画像をどのような画像でシーンに分割するかを選択する。

【0075】例えば、スキージャンプ競技の映像に対して、各選手のジャンプごとにシーンに分割したいと考えているとする。このような場合には、各選手のスタート場面の映像を一つ選択しておく。類似画像検索部107は選択された画像と他のフレームとの類似度を計算する。類似度の計算方法としては、最も簡単な例としては、二つの画像の対応する位置の輝度や色の差の総和を求める方法、二つの画像の輝度または色のヒストグラムの差分和を求める方法などがある。このとき、利用者が類似度の計算方法のパラメータを入力できるようにして置いても良い。例えば、スキージャンプのスタートのシーンでは、背景はほとんど変わらずに選手の服の色、柄だけが主に変わるので、選手の移っている領域は類似度の計算に含めないように指定することができる。

【0076】利用者は動画画像処理装置に行わせる処理を指示する際、キーボードやマウス等の入力装置を用いるが、これは図13には省略してある。図14で説明され

るように、フレーム201を選択してシーンの分割を行うと、類似画像としてフレーム205、フレーム206等が類似画像検索部107により検出される。シーン情報生成部106は、類似していると判定されたフレーム、またはこれらフレームを含むショットを境にして複数のショットをシーンに分ける。例えば、二つの類似したフレーム201、205の間のショット207、208、209、210が同一のシーンであると判定する。ショット207からショット210までが同一のシーンであるという情報は動画データとともに記憶部104に蓄積される。また、類似画像検索のキーとなったフレームと、検索された類似画像はそれぞれのフレームの含まれているシーンの代表フレームとしてシーン情報に登録される。シーン情報は実際にはシーンのはじめ、終わりのタイムコード、フレームナンバー、ショットナンバー、記憶アドレス、また、フレーム数やシーンの長さの全てまたは一部によって記述される。

【0077】また、類似した画像であるフレーム201、フレーム205、フレーム206等には、リンク生成部108により類似画像であるという関連が定義される。すなわち、これらのフレーム（もしくはフレームを含むショット）の間にリンクが張られる。リンク情報もシーン情報と同様、関連づけられたフレームのタイムコード、フレームナンバー、ショットナンバー、記憶アドレスなどの一部または全てにより記述される。

【0078】検索の際にはシーン情報、リンク情報を利用してシーン単位での動画の一覧ができる。例えば、利用者がシーン単位の一覧表示を行うよう指示を入力すると、それぞれのシーンの代表フレームが図15のように一覧表示される。このとき、各シーンを代表するフレームは、動画データとともに蓄積されているシーン情報、リンク情報を読み出し、はじめのシーンの代表フレームおよびこのフレームに関連の定義されているフレームを次々に表示させていくことにより実現する事ができる。

【0079】利用者が図15で表示されている代表フレームのどれかを指示すると、その代表フレームに対応するシーンの再生が開始される。所望の場面がどの代表フレームと対応しているかを確認する際や、実際に所望の場面を見る場合にこのような操作を行う。

【0080】また、シーンの中に含まれるより短い場面を検索しているような場合には、シーン単位の代表フレーム一覧表示ではなく、あるシーンの中のショット単位の代表フレーム一覧表示に切り替えることができる。そのためには、利用者はショット単位の一覧表示を行う旨の指示と、ショット単位の表示を行うシーンの指定とを装置に対して行う。このような指示が行われると、動画データとともに蓄積されているカット検出の結果情報を読み出し、カットにより区切られるショットごとに代表フレームを一覧表示させる。ショットの代表フレーム

はあらかじめ利用者によって登録されたものや、単にショットの先頭のフレームを選ぶ方法などが選択できる。図16は図15で表示されたシーンの代表フレームのうちの一つを選択し、対応するシーンの中のショットの一覧表示を行ったときの表示例である。表示されているショットの代表フレームを選択すると、そのショットに対応する場面からの再生が開始される。

【0081】以上ではシーンの代表フレーム一覧からシーンを選択し、そのシーンに含まれるショットの代表フレームを一覧させていたが、逆にショットの一覧を行った後、シーンの一覧に切り替えることもできる。

【0082】例えば、はじめに全てのショットの代表フレームの表示を行ったが、表示されたフレームの数が多すぎるといった場合に、フレームの表示への変更を指示することができる。

【0083】スキージャンプ競技の映像においては、選手のスタートシーンを類似画像検索のキー映像とすることにより動画全体を選手ごとの競技シーンに分割することができた。本実施例で説明される発明は、このように類似したシーンが繰り返し含まれ、しかもその一回の繰り返しを一つのシーンとしたいときに特に効果がある。相撲のたちあい場面をキーにしてシーンを区切れば一つ一つの取り組みがシーンとして分割されるし、短いTV番組をいくつか連続して一つの動画データとしている場合には、番組の始めの場面をキーとして分割することにより、シーンと一回の番組が対応する。

【0084】また、シーン分割後に数シーンに一度しか出てこないような場面を類似画像検索のキー映像として選択し、複数のシーンをまとめるような分割を行い、さらに階層化されたシーンの構造を作ることにも可能である。このとき、一覧表示されたシーンの代表フレームのうちの一つを選択すると、そのシーンに含まれる一つ下の階層のシーンの代表フレームの一覧表示が行える。このように、動画の内容、長さに応じて動画をシーンに分割し、効率よく所望の場面を検索することができる。

【0085】以前に見た動画の中から所望の場面を検索する場合だけでなく、初めて見る動画の中から興味のあるような場面を選択する場合にも、上記のような一覧表示は有効である。ショットの一覧表示だけでなくシーンという一つ上の階層での表示ができるため、利用者が動画の内容を理解しやすいフレーム数で一覧表示することができる。

【0086】（第二の実施例2）次に、本発明の第二の実施例2を説明する。本発明の第二の実施例1では、類似画像検索部107で検索のキーとする映像は利用者により指定されていた。本実施例では、利用者がわざわざキーとなる映像を指定することなく、動画をシーンに分割し、この分割情報を元に所望の場面を効率よく検索する方法について説明する。

【0087】記憶部104に記憶された動画データは、すでにカット検出部102によりカットの検出が行われ、カット情報が付与されているものとする。利用者によるキー映像の選択が行われない場合には、類似画像検索部107はカット情報により分割された動画の各ショットごとにキー映像となるフレームを順次選択していく。例えば、各ショットの先頭のフレームを順次キー映像として選択していく。類似画像検索のキーとなる映像が選択された後の処理は、第一の実施例で説明したのと同様である。類似していると判定されたフレームで区切られるショットの集合はシーン情報生成部106によりそれぞれシーンとしてまとめられ、シーン情報として動画データとともに記憶部104に記憶される。類似したフレームの間にはリンク生成部108により関連が定義され、リンク情報は記憶部104に記憶される。

【0088】一つのキー映像に対し、類似画像の検索、フレーム情報の生成・蓄積、リンク情報の生成・蓄積が終了すると、次のキー映像に対して同様の処理を行う。このとき、図17で説明されるように、キー映像をどの映像にしたかにより、複数の種類のシーンが生成される。図17中、303はフレーム301の類似画像に張られた類似画像を示すリンク、304はフレーム302に張られた類似画像のリンクである。また、305はフレーム301をキー映像としてシーン分割したときの一つのシーンを表し、306はフレーム302をキーとしたときの一つのシーンを表す。このように、それぞれのシーン情報はキー映像と関連づけて記録されねばならない。

【0089】類似画像検索部107は全てのショットから選択されたキー映像を検索キーとして類似画像検索を行う必要はない。動画のはじめのショットから順次キー映像を選んでいき、途中キー映像として選ばれたショットに類似した画像が存在する旨のリンクがすでに張られている場合には、これをキーとして類似画像の検索およびシーンへの分割を行う必要はない。それまでに行ったシーン分割の中に同じ結果が得られているはずだからである。

【0090】以上の手続きによって、複数のシーンへの分割情報が動画データとともに蓄積されることになるが、これを利用した所望の場面検索を行うためには、まず類似画像検索を行った際のキー映像の選択を行う必要がある。このためには、利用者はキー映像の一覧表示を行う旨の指示を入力してキー映像一覧表示を行い、その中から望みのキー映像を選択すればよい。例えば、図15はスキージャンプのスタート時のフレームをキーとして選択したときのシーン一覧であるが、飛躍中のフレームをキーとして選択したときには各選手の飛躍中のフレームが代表として一覧表示されることになる（ただし、代表フレームはシーンのはじめのフレームが選ばれているとき）。このように、好みのフレームをキー映像とし

て選択することにより、異なったシーンの分割、表示が行える。

【0091】（第二の実施例3）以下、本発明の実施例を図を用いて詳細に説明する。第二の実施例1、2では類似したフレームによって区切られるショットの集まりをシーンとして一つにまとめていた。本発明の第二の実施例3はこれとは異なり、類似したショットを同一のシーンとしてまとめていく。第二の実施例3を実現するための構成ブロック図はやはり図13で与えられる。しかしながら、シーン情報生成部106及び類似画像検索部107の働きが第二実施例1、2とは異なっている。

【0092】本実施例においては、類似画像検索部107は動画のショットとショットの間の類似度を計算する。このとき、比較される二つのショットはあらかじめ定められたショット間隔以内のもの、もしくは、あらかじめ定められた時間間隔以内のものに限定される。上記の二つのショットの類似度計算は、定められたショット間隔以内に存在する、もしくは定められた時間間隔以内に存在する全てのショットの組み合わせについて全て求める必要はなく、どこかしらのシーンに含まれたショットは次々に対象から取り除いていっても良い。このようにした方が計算時間を少なくすませることができる。類似度は、例えば、同一のショットに含まれるフレームの平均画像を求め、二つの平均画像の相関値や差分を求める方法、それぞれのショット内のフレームの色ヒストグラムの差分値の和をとる方法などがとられる。相関値があらかじめ定められたしきい値よりも大きなショット同士は、同一のシーンに含まれるものとされる、シーン情報生成部106は類似画像検索部107からの類似したフレームの情報を受け、類似したショットが同一のシーンである旨のシーン情報を生成し、動画データとともに記録装置104に記録する。シーン情報は具体的にはシーンナンバー、シーンに属するショットナンバー、もしくはフレームナンバー、アドレス、タイムコード等によりなる。

【0093】例えば、室内で撮影されたシーン（ショットの集まり）に続いて屋外で撮影されたシーンが続いている場合、明るさなどの映像の特徴がどちらのシーンであるかにより変わっている。従って、適当に類似度のしきい値を設定することにより、ショットがどちらのシーンに属しているのかを判定することができる。図18はシーン分割の様態を示した説明図である。ショット401をキーとして計算した他のショットとの類似度が407である。ショット402～405は類似度がしきい値408を越えているので同一シーンと判定されるが、ショット406以降は類似度がしきい値よりも低いため、同一のシーンではないと判定される。

【0094】シーン情報が生成された後は、第二の実施例1と同様に、シーンごとに代表フレームを表示させ、利用者がこれら表示されたフレームの中から一つを選択

することにより、動画の再生やシーン内のショット一覧表示を行うことができる。

【0095】（第二の実施例4）本発明の第二の実施例4は、図19のブロック図のごとく構成することにより実現される。文字・シンボル検出部501は、動画像中に含まれる文字・シンボルの検出を行う。検出の手法としては、例えば情報処理学会第48回前期全国大会1M-5や同1T-7に報告されている手法が適用される。検出は、動画像を映像取り込み部101で取り込みながら行う場合と、すでに記憶部104に取り込まれた動画像データに対して行う場合とがある。文字またはシンボルが含まれていると判定された場合には、フレーム情報、文字列が認識可能である場合には認識された文字列情報、また、認識が不可能な場合には単に文字列が含まれているという情報、及びフレーム中の文字・シンボルの位置、フレームナンバーなどの文字・シンボル情報が動画像データとともに記憶部104に記憶される。

【0096】文字画像検索部503は動画像の中から特定の文字・シンボルに関する条件を満たすフレームを検出する。また、代表フレーム情報生成部502は文字画像検索部503から送られてくる検索結果から動画像を代表するフレームに関する情報を生成し、動画像データとともに記憶部104に記憶させる。以下、これらの処理の流れを詳細に説明する。

【0097】例として、ニュース番組の映像が動画像データとして記憶部104に取り込まれているものとする。この動画像データはすでにカット検出部101によりカット情報が付与されている。さらに、文字・シンボル検出部501により文字・シンボルの含まれているフレームが検出され、文字・シンボル情報が動画像データとともに保持されている。利用者が動画像中の所望の場面を探す際には、カットとカットとにより分割されたショットごとに代表フレームをあらかじめ定められた方法で選び（例えばショットの先頭のフレームを選ぶ方法や、第五の実施例にhで説明される方法）、一覧表示させ、その中から選択することができる。しかしながら、一動画中のショット数は大変多いため、これらを一覧表示する場合には何度かに分けて表示したり、代表フレームを小さくして表示しなければならない。

【0098】このような場合に、本実施例の動画像処理装置では、まずすでに選ばれている代表フレームのうち文字情報が含まれているフレームだけを文字画像検索部503が選択する。これは動画像データとともに記録されている文字・シンボル情報を参照することにより、容易に選択できる。そして、選択された代表フレームだけを一覧表示する。ニュース番組では、ニュースの話題が変わったときにテロップ等の文字情報として動画中にインポーズされていることが多い。

【0099】そのため、代表フレームの中から文字情報を含むフレームだけを代表として表示させると、一度に

表示される代表フレームの数を少なくすることができるだけでなく、内容のわかりやすいフレームを選択した一覧表示が可能である。このとき、文字を認識することが可能であったならば、代表フレームではなく認識された文字列を一覧表示して利用者に選択させる方法をとることもできる。

【0100】一覧表示したフレームもしくは文字列が利用者により選択されると、対応する場面から動画像を再生する。もしくは、選択されたフレームもしくは文字列とその次に表示されているフレームもしくは文字列との間に存在している代表フレームの表示を行う。はじめに、文字列を含む代表フレームのみの表示で大まかに場面を選択しておき、次に代表フレームの表示で細かに場面を指定する事により、効率よく所望の場面を検索できる。

【0101】文字画像検索部503には利用者の指示により、動画像に含まれている文字・シンボル情報のうち、所望のものだけを検索する昨日を持たせることができる。例えば、文字・シンボルの出現位置や種類を指定することができる。例えば、ボクシングの動画像において画面右下の経過時間の表示が0であるフレームだけを表示させれば、各ラウンドごとに先頭の一枚の代表フレームだけが表示され、少ない枚数で整理されたフレームの表示ができる。

【0102】また、野球のゲームの動画像において、ノーアウト（さらにノーストライク、ノーボール）を示すシンボルを含むフレームを検索し、一覧表示させれば、各回の表、裏の攻撃の始まるシーンだけが選択されて表示されることになる。画面下に多くの文字情報が出されているフレームを選択して表示させれば、打者情報を表示させている場面のフレームが多く表示されることになり、実質的には打者ごとに代表フレームを表示させていることになる。

【0103】文字の認識が可能な場合には認識された文字情報を使い、同一の文字列を含むフレームのみの選択、これらのフレームの間の関連を定義することができる。例えば、野球において好みの打者名が表示されているフレームのみの一覧表示を行えば、好みの打者の打席の中から好みのものを即座に選択することができるし、利用者が関連をもつフレームへのジャンプ命令を入力すれば、リンク情報から関連フレームのタイムコード、記録アドレス等の情報を読み出し、即座に次の打席の場面の再生が行われる。

【0104】動画中にインポーズされた文字の認識が可能で、かつ辞書情報が利用可能な場合には、認識された文字列に対応する辞書を参照することにより関連語をピックアップしておき、他の場面で認識された文字列が関連語に一致する場合にも二つの場面にリンクを定義することができる。辞書の利用により、より多様な意味づけを持ったリンクを自動生成することができる。

【0105】（第二の実施例5）次に、第二の実施例5として、ショットの代表フレームを選択する方法について説明する。最も簡単なショットの代表フレームの選択方法は、ショットの先頭のフレームを選択する方法である。しかし、この方法では必ずしもショットの内容を反映したフレームが選択されるとは限らない。

【0106】本発明によるショットの代表フレームの選択方法の処理の流れを図20に示す。まず、代表フレームがまだ選択されていないショットの中から一つのショットを選ぶ。そのショットの中のフレーム全てに対して、あらかじめ利用者により選択されたキーフレームとの類似度を求め、類似度が最大のフレームをそのショットの代表フレームの候補とする。類似度の計算方法としては、すでに述べたように、二つの画像の対応する位置の輝度や色の差の総和を求める方法、二つの画像の輝度または色のヒストグラムの差分和を求める方法などが用いられる。候補となったフレームの類似度があらかじめ定められたしきい値以下の値しか持たない場合には、ショットの先頭（もしくは中間のフレームなど、定められた位置）のフレームを代表フレームとして選択する。

【0107】例えば、人の顔のアップの画像をなるべく代表フレームとしたい場合には、人の顔のアップの位置フレームをキーフレームとして選択しておく。この場合、画面の中心に大きく肌色の領域が存在し、上部もしくは左右に頭髮の色が分布している。

【0108】このようなフレームをキーフレームとして代表フレームの選択をさせれば、類似したフレームとしてやはり顔のアップが多く選ばれてくる。顔のアップがないショットでは、あらかじめ定められたショット内の位置のフレームが選択されてくる。

【0109】このような代表フレームの選択方法は、ショットの代表を選ぶ場合に限られるものではない。動画画像がショット以外の何らかの単位でフレーム群に分割されている場合、一つのフレーム群から代表フレームを選択するために全く同じ手法を用いることが可能である。

【0110】本発明の別のショット代表フレーム選択方法は、第二の実施例4のごとく、文字・シンボル情報を含むフレームをショットの代表フレームとして選択する方法である。図21はこの処理の流れを表したものである。

【0111】まず、代表フレームがまだ選択されていないショットの中から一つのショットを選ぶ。動画画像データとともに記録されている文字・シンボルデータを参照し、ショット内のフレームに文字・シンボルを含むフレームが存在するかどうかを調べる。存在しなければ、ショットの先頭の（もしくは中間などショット内の決められた位置の）フレームを代表フレームとする。もしも文字・シンボルを含んでいるフレームが存在するならば、それらのうちの先頭のフレームを代表フレームとする。

【0112】すでに説明したように、文字情報はショッ

トの内容を理解するための情報を多く含んでいる。例えば、ニュース番組において、単にキャスターが移っているフレームが代表フレームとして選択されていても、そのショットでどのような話題がはなされているのかはわからない。しかし、ニュースの話題がテロップとしてインポーズされているフレームを代表フレームとして選択しておけば、代表フレームを見るだけでそのショットの内容を理解することができる。

【0113】（第二の実施例の効果）本発明の動画画像処理装置では、動画画像が細かく分割されているときに、これら分割された部分のいくつかをまとめてより大きく分割された動画画像を構成し、それぞれの部分について一つの代表フレーム、もしくはそれに関する情報を表示することができるので、表示される情報が多くなりすぎることがなく、利用者が効率的に所望の場面を検索することができる。

【0114】さらに、利用者がこれら分割を自ら行ったり、代表フレームもしくはそれに関わる情報を一つ一つ選択していくことが不要であるため、利用者の手を煩わせることがない。

【0115】（第三の実施例）以下、本発明の第三の実施例を図面に基づいて説明する。混乱をさけるために、以下では撮影を中断せずに撮影した一連の動画画像を「ショット」と称し、動画画像の構成、内容の連続性などから複数のショットを統合したものを「シーン」と称することにする。また、連続する2ショットの境界を「カット」とし、同様にシーンの境界を「シーンチェンジ」と称する。

【0116】図22は本発明の一実施例に係わる映像データベースの表示を示す概念図である。1101は動画画像のある時間区間を表す立体アイコンである。立体アイコン1101の最前面1102は、通常、動画画像における、その時間区間の開始時刻での映像であるが、ショット内の特徴的なフレームを自動検出あるいは利用者の入力により指定して表示してもよい（代表フレームの選択方法については参考文献：「内容把握を支援する動画インデックス」1993年電子情報通信学会春季大会講演予稿集 SD-9-4など）。

【0117】また、立体アイコン1101の奥行きは、その時間区間の長さに対応しており、利用者は立体アイコンの奥行きの深浅で、その区間の継続時間を知ることができる。従来技術では、この立体アイコンはショット毎に表示されていたが（ショットの分割方法については例えば参考文献：「テレビジョン画像のシーン境界の検出法に関する一考察」電子情報通信学会技術報告 IE94-49など）、本発明の表示方式においては複数のショットを統合して一つの立体アイコンとして表示することが特徴の一つである。この統合は、一覧表字全体の見やすさを考慮して、利用者自身が統合の度合を可変的に決定できるようにしていてもよい。

【0118】この場合、例えばある動画像が10シーン40カットである場合には、そのままシーンごとの立体アイコンとして表示し、50シーン200カットの場合にはシーンごとの関連の強さを計算して、さらにシーンごとの統合を進め、立体アイコンを20程度にまで統合して表示する、などである。複数のショット（あるいはシーン）の統合方法については、ショット内のフレームについて色調の平均値やヒストグラムを求めて、ショットどうしで比較する方法や、人物の認識を行って、登場の分布から連続性を判断する方法などが考えられる（参考文献：「色特徴による映像シーン分類の検討」1993年電子情報通信学会春季大会講演予稿集 D-490）。

【0119】立体アイコン1101は2次元座標に投影されてコンピュータなどの画面に表示される。耳1103は元のカット毎に付けられ、奥行き方向に埋もれてしまったショットの存在を明示する。耳1103には文字情報を表示することができ、利用者が後の編集のためにそのショットに対するメモ文を入力すると、そのメモ文が記憶され、表示される。未入力のショットに対しては、そのショットの時刻を表示するなどする。耳1103上の文字情報は表示しなくてもよい。

【0120】立体アイコン1101は最前面1102に直交する面（空間的には上面、底面、右側面、左側面の4面をもつ。以下ではこの4面を総称して単に「アイコン側面」という。

【0121】図22では右側面1104と上面1105を図示。）をもつが、アイコン側面には、立体アイコンの断面が投影されて表示されている。これについて図24を用いて説明する。

【0122】ある動画像の時刻 t における映像3301を、画素 (x, y) のRGB値を用いて $(r(x, y, t), g(x, y, t), b(x, y, t))$ と表わしたとする。

【0123】このとき、 xyt 空間の、時間成分を持つ平面3302は $ax + by + ct = d$ と表わされる。この平面3302による映像3301の断面を上面3303に投影するときは、座標 (x, t) にRGB値 $(r(x, (d - ax - ct)/b, t), g(x, (d - ax - ct)/b, t), b(x, (d - ax - ct)/b, t))$ を描画することで投影図が決定される。

【0124】同様に右側面3304に投影するときには、座標 (y, t) にRGB値 $(r((d - by - ct)/a, y, t), g((d - by - ct)/a, y, t), b((d - by - ct)/a, y, t))$ を描画すればよい。

【0125】動画像の大きさが幅 W 、高さ H で、画面の右上隅を $(x=0, y=0)$ とする場合、平面3302の選び方としては、上面に対して「 $y=0$ 。 x, t は任意」、右側面に対して「 $x=W$ 。 y, t は任意」、底面

に対して「 $y=H$ 。 x, t は任意」、左側面に対して「 $x=0$ 。 y, t は任意」と設定するのが分かりやすいと想定される。

【0126】この場合、各アイコン側面に投影される画像は、動画像の時刻ごとの画像をフチなしの写真にプリントし、時間順に複数枚重ねて横から見たときの様な画像になる。このような側面表示をおこなうことにより、カメラの左右の移動（パン）を上面に見ることができるほか、たとえば右からオブジェクトが侵入してきた場面においては右側面にオブジェクトの姿が見い出せるなど、ショットの内容把握に効果がある。

【0127】図22にもどって説明を続ける。立体アイコン1101の近傍にはオブジェクト1106や文字1107が表示されている。これらは、参照線1108などを媒介して、それが登場した時刻のアイコン側面に対応づけられて表示される。

【0128】表示すべきオブジェクトまたは文字は、人物認識技術や文字判定技術によって自動認識で判断させてもよいし（オブジェクトの抽出については参考文献：「配色情報に基づく画像からのオブジェクトの抽出と検索の検討」1994年電子情報通信学会春季大会講演予稿集、D-422など。文字領域の切り出しについては参考文献：「階層的な色分類によるカラー画像からの文字パターン抽出法」1994年電子情報通信学会秋季大会講演予稿集、D-289など）、利用者があらかじめ画面を見ながら選択、決定しておいてもよい。

【0129】いずれの方法においても、そのオブジェクトまたは文字がある時間にまたがって画面内に存在する場合、そのオブジェクトまたは文字の領域が最大の面積になった時刻のオブジェクトまたは文字を1106または1107のように表示する。このようにオブジェクト1106や文字1107を表示する対象は、必ずしも複数のカットが統合された立体アイコンに対してとは限定されず、一つ一つのカットを独立の立体アイコンとして表示している場合にも表示できる。なお、これらオブジェクト1106や文字1107の表示については、利用者によって表示、非表示の状態を選択してもよい。

【0130】奥行き表示は必ずしも立体アイコン1101の一部でなくてもよい。例えば図28のように、元の映像情報の1フレームを表示しない側面のみを表示し、耳1103やオブジェクト1106A～Cなどを表示してもよい。また、図28では原画像の断面ではなく、各オブジェクトの存在位置を図示した表示を示している

（表示平面6601上で、奥行き方向、すなわち図28の左下から右上が時間の進行方向）。このように、「該映像情報の一部」とは、映像×時間の3次元を平面で切った断面の写影図に限られず、時刻時刻でオブジェクトがどの位置に存在していたかなどの副次的な情報を意味してもよい。

【0131】図28の表示は、上記のような手段で動い

ているオブジェクトや人物を認識抽出したあと、そのオブジェクト領域の重心が、画面の左右座標のどこに位置したかを計算した結果である。オブジェクト1106Aに対応する動き表示は6602A、以下同様にB、Cと対応している。ここで各AとCは同じオブジェクト（登場人物）を指しているが、システム内部では、必ずしもAとCが同じオブジェクトであることを認識していなくてもよい。この場合、利用者は表示平面6601から引き出されたオブジェクト1106A～Cを見ることにより、目視でAとCの一致を確認し、これらが同一であるという定義をシステムに対して行うことができる。

【0132】図28のような表示を提供することによって、より多くのオブジェクトが存在する場面を利用者が優先的に検索したり、目的の登場人物が中央に写っている場面を検索したりする際の効率がよくなる。

【0133】つぎに、立体アイコン1101上を指示することによる映像の提示方法について説明する。カーソル1109は立体アイコン1101と同じ画面上にあり、マウス、トラックボールなどのポインティングデバイスによって利用者の指示したい位置を示している。カーソル1109で立体アイコン1101の右上辺1110上の一点を指示することによって、対応する映像時刻は指示点の奥行き座標から一意に決まる。すると、その時刻の映像（静止画）が上面1105上に引き出されるように表示される（画面1111は引き出された状態を示す）。

【0134】引き出された画面1111は、カーソル1109が引き出された位置から去ると消える。あるいは引き出した状態でマウスボタンをクリックするなどの操作を行うことによって、消去の指示があるまで画面に残留してもよい。右上辺1110上の一点を正確に指示することは難しいので、右上辺1110と垂直方向に数画素の範囲に指示があった場合でも、その指示位置から右上辺1110上に垂直に下ろした点を指示した場合と同様の動作をしてもよい。

【0135】図25と図27には、図22のような表示が画面上になされているとき、指示位置によって発生する動作の例を示す。「そのショットを選択」「そのシーンを選択」とは、動画像の編集のための動作であり、ショットやシーンを複数選択して、ショットやシーンをつなぎあわせたり、画像効果を与えたりする際に用いる。

【0136】一方、カーソル1109による逐次指示を与えなくても、動画像再生の指示を与えることで自動的に画像が立体アイコン1101から次々と引き出されていってもよい。また、すでに再生された画像部分に関しては立体アイコンから削除され、再表示の指示があるまで画面に現われなくてもよい。この場合、たとえば順次画像を再生してゆくに従い立体アイコン1101は短くなり、最前面1102に表示される画像も時間進行してゆく。このように最前面で時間進行が再現できる場合に

は画面1111のように引き出した表示をする必要はない。

【0137】また、引き出し表示の代わりに別の表示画面（図示せず）をもち、そこに指示した時刻の画像を表示してもよい。たとえば表示画面を立体アイコン1101が表示されている画面と別の画面に設け、立体アイコン1101のある画面を操作者が見、指示時刻の表示画面を操作者と異なる人物（客）に見せれば、操作の様子を客に見せずに画像提示ができるため、動画像を効果的にプレゼンテーションすることができる。

【0138】図27は、図22のような表示を行うための処理手順を例示したフローチャートである。システムはまず、フレームを読み込むことができるかどうかを調べる（S501）。読み込むことができないければ、画面作成は終了する。読み込むことができるとき、そのフレームの画像データおよび画像の属性（カット／シーンチェンジか、など）を読み込む（S502）。そのフレームがシーンの先頭であった場合（S503）、これまで奥行き方向に積み上げてきた立体アイコンの更新をやめ、これまでに作成した立体アイコンの隣などに次の立体アイコンを作成するように準備する（S504）。そして立体アイコンの最前面にそのフレームの画像を表示する（S505）。シーンの先頭は一般にショットの先頭でもあるので、耳をつける（S506）。

【0139】一方、そのフレームがシーンの先頭でなくてもショットの先頭であった場合（S507）、立体アイコンのアイコン側面にショットの区切りを示す区切線を書き込み（S508）、耳をつける（S506）。

【0140】次に（シーンやショットの先頭であった場合でも、そうでなかった場合でも）、オブジェクトや文字を表示すべきかどうかを判定する。上述したように、オブジェクトや文字は、それらがそのショット内で最大であった時、表示するのが適切なので、同一のオブジェクトや文字がフレームの進行とともに大きさや形、位置をどう変化させているのかを、システムは追跡している。そしてそれまで追跡してきたオブジェクトや文字の大きさが拡大から縮小に転じたときに、オブジェクトや文字を立体アイコンの側面近傍に表示する（S509、S510、S511）。このとき、文字認識技術を用いて抽出した文字を認識し、自動的にそのショットのメモに利用してもよい（耳に書き込んでもよい）。

【0141】また、オブジェクトや文字の大きさが正確に最大であった箇所一点で表示を行いたい場合には、それぞれのオブジェクトについて、最大になった時刻と大きさ、領域を記憶しておき、ショットの終端が来たときに記憶していた時刻にさかのぼって表示させてもよい。

【0142】次に、1フレーム分だけ立体アイコンの厚みを増やし（S512）、側面の画像についても、そのフレームに関して書き加える（S513）。以上の過程を処理残りのフレームがなくなるまで繰り返す。

【0143】以上の説明では、立体アイコン1101の最前面1102の画像や、立体アイコン1101の近傍に表示されたオブジェクト1106などは静止画像としてきたが、これらは動画像として表示してもよい。例えば最前面1102の画像であれば、その立体アイコン1101の範囲のシーンを繰り返し再生していてもよい。

【0144】また、その立体アイコン1101に含むカットの画像を数秒ごとに交互に表示してもよい。あるいは、オブジェクト1106については、そのカット内でのオブジェクトの動作を表示するなどすることができる。

【0145】このような立体アイコン最前面の動画像表示は図22のように複数のカットを統合表示した場合に限られず、図23のようなショット列挙の場合に用いてもよい。この場合は各立体アイコンの最前面2201に、そのショットの画像を繰り返し再生するなどする。

【0146】（第三の実施例の効果）以上説明したように本発明によれば、動画像編集や視聴の際の一覧提示を、効率的に統括することができ、長い動画像やショットの多い動画像を扱う際にも利用者の圧迫感を軽減した操作環境を提供することができる。

【0147】

【発明の効果】以上説明したように本発明の動画像処理装置によれば、長時間に及ぶ動画像を比較的軽微な計算処理によって時分割（カット）することができ、更にそのカットは画像内容の変化点であるから、意味情報を反映した分割であることが期待できる。

【0148】一方、本発明の動画像処理装置では上記手法によって分割されたカットを統括し表示する機能を持つので、利用者にとってもわかりやすい操作で効率よく目的の場面に到達することができるような動画像編集・検索環境を提供できる。

【図面の簡単な説明】

【図1】本願発明の第一の実施例1の一構成例を示すブロック図である。

【図2】本願発明における映像カットの検出部の一構成例を示すブロック図である。

【図3】本願発明の動きベクトルブロックの説明図である。

【図4】本願発明の映像カットのない場合の修正動きベクトルの例の図である。

【図5】本願発明映像カットのある場合の修正動きベクトルの例の図である。

【図6】本願発明の動き補償を使った画像圧縮におけるフレームと検出される動きベクトルの関係図である。

【図7】本願発明の順方向の動きベクトルと順方向の修正動きベクトルの関係図である。

【図8】本願発明の逆方向の動きベクトルと逆方向の修正動きベクトルの関係図である。

【図9】本願発明の映像変化量としきい値を使った映像

カット判定の例の図である。

【図10】本願発明の第一の実施例2の一構成例を示すブロック図である。

【図11】本願発明の第一の実施例3における映像カット検出部の一構成例を示すブロック図である。

【図12】本願発明の第一の実施例の動作の手順を示す図である。

【図13】本願発明の第二の実施例1, 2, 3の構成を表すブロック図である。

【図14】本願発明の動画像のリンク定義とシーン分割の説明図である。

【図15】本願発明のシーナー一覧表示の例の図である。

【図16】本願発明のショット一覧表示の例の図である。

【図17】本願発明の自動シーン分割の説明図である。

【図18】本願発明の第二の実施例3におけるシーン構成方法の説明図である。

【図19】本願発明の第二の実施例4の構成を表すブロック図である。

【図20】本願発明の第二の実施例5における処理の説明図である。

【図21】本願発明の第二の実施例5における処理の説明図である。

【図22】本願発明の映像データベースの表示方法に係る概念図である。

【図23】従来の映像データベースの表示方法に係る概念図である。

【図24】本願発明の映像データベースの表示方法を説明するための概念図である。

【図25】本願発明の映像データベースの表示方法の利用例を示す概念図である。

【図26】本願発明の映像データベースの表示方法の利用例を示す図である。

【図27】本願発明の映像データベースの表示方法を実現する処理例を示すフローチャートの図である。

【図28】本願発明の映像データベースの表示方法の利用例を示す図である。

【符号の説明】

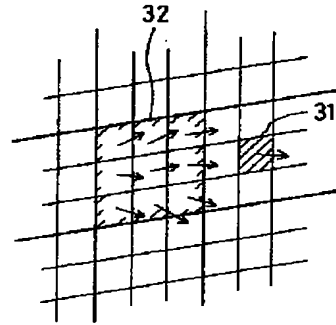
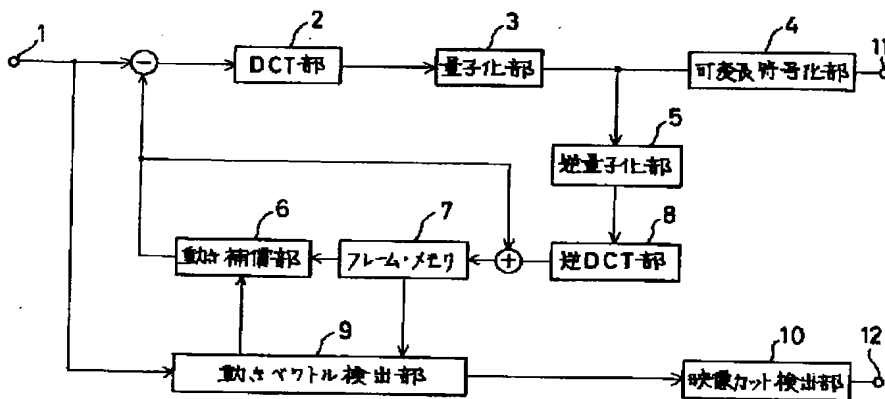
1 映像入力端子	10 映像カット検出部
2 DCT部	11 圧縮映像出力端子
3 量子化部	12 映像カット情報出力端子
4 可変長符号化部	21 動きベクトル修正部
5 逆量子化部	22 映像変化量算出部
6 動き補償部	23 比較部
7 フレーム・メモリ	24 映像変化量メモリ
8 逆DCT部	25 しきい値算出部
9 動きベクトル検出部	
100 動画像データ入力端子	

101 映像取り込み部
102 カット検出部
103 表示部
104 記憶部
105 表示画像検索部
106 シーン情報生成部
107 類似画像検索部

108 リンク生成部
1101 立体アイコン
1103 ショットを示す耳
1106 抽出されたオブジェクト
1107 抽出された文字
1109 カーソル
1111 引き出し表示された画面

【図1】

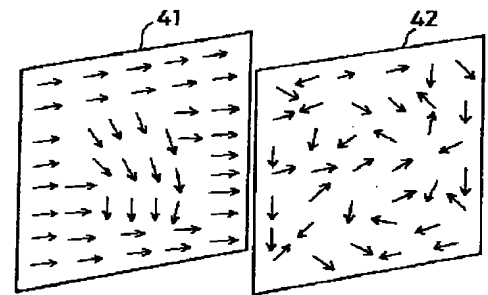
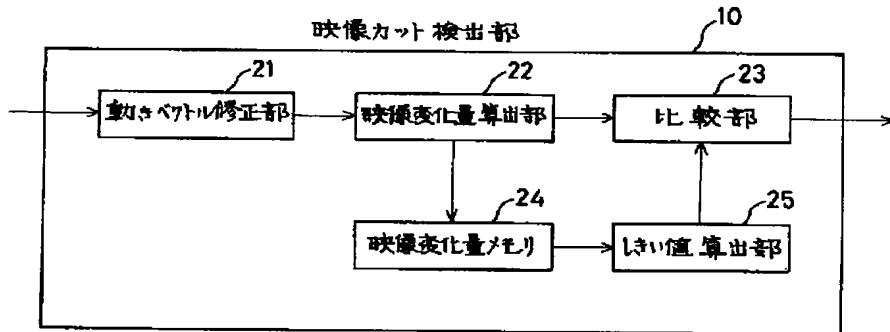
【図3】



【図4】

【図5】

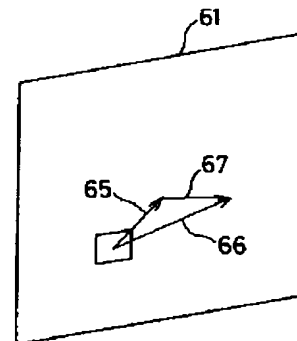
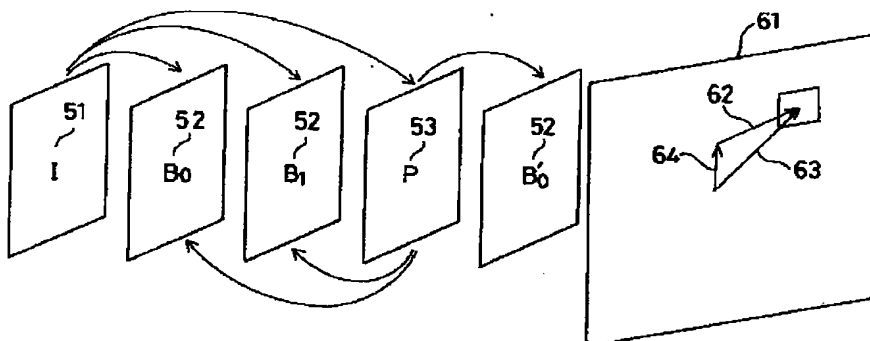
【図2】



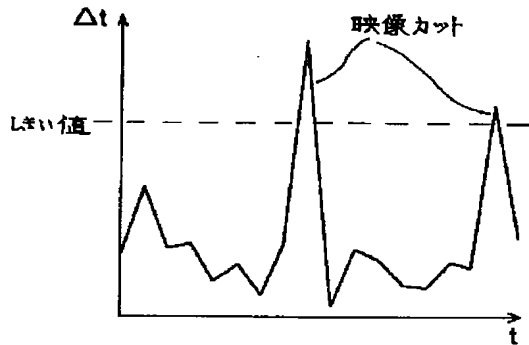
【図8】

【図6】

【図7】



【図9】

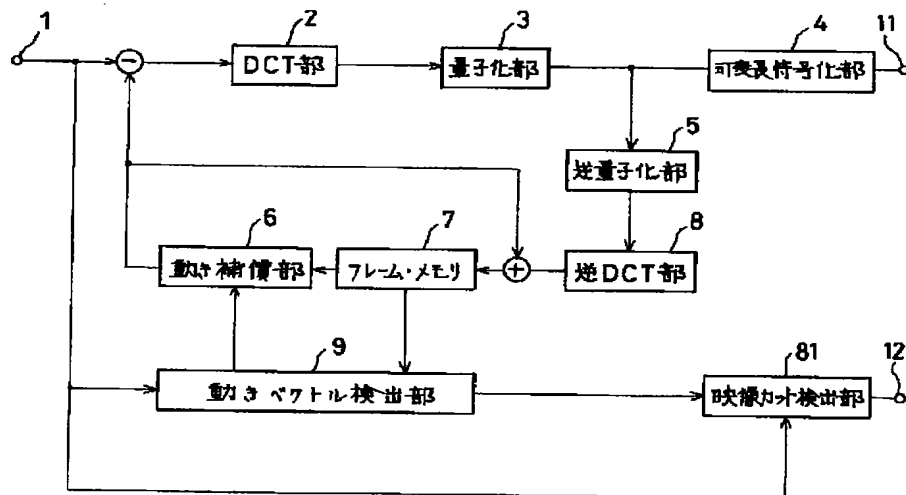


【図26】

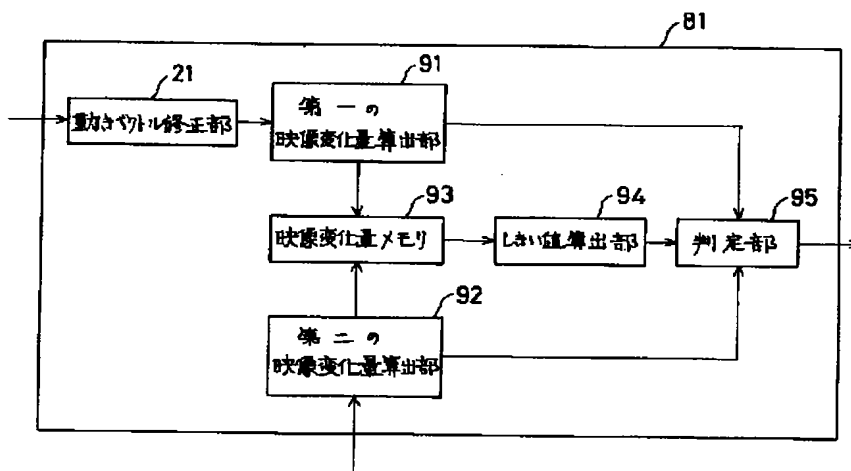
図4上の各領域を指示したときの動作

領域	動作
A	それが差し示すフレームを表示(引き出す)
B	そのショットを選択
C	位置に対応した時間のフレームを表示(引き出す)
D	引き出し画面を消去
E	そのシーン(立体アイコン全体)を選択

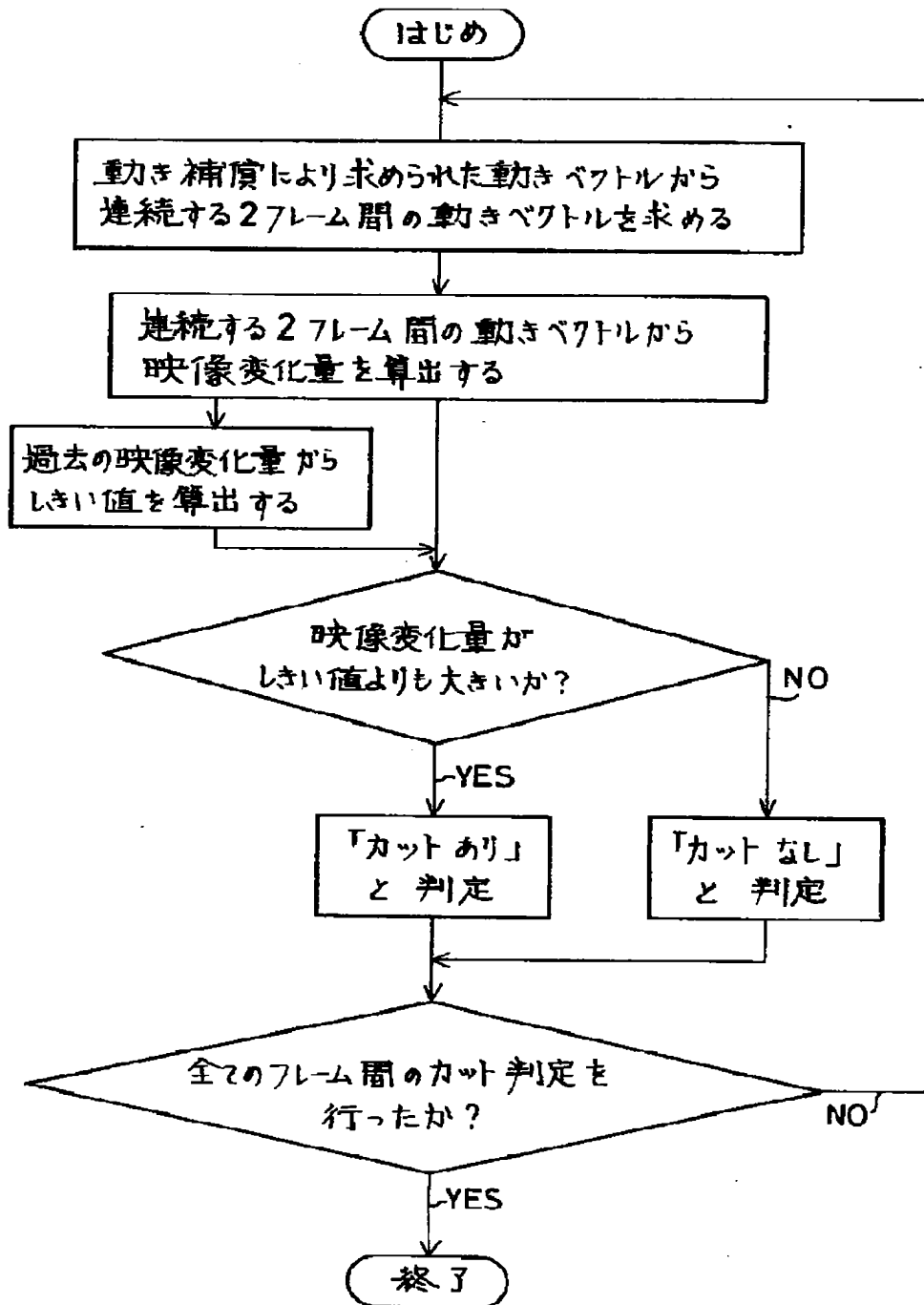
【図10】



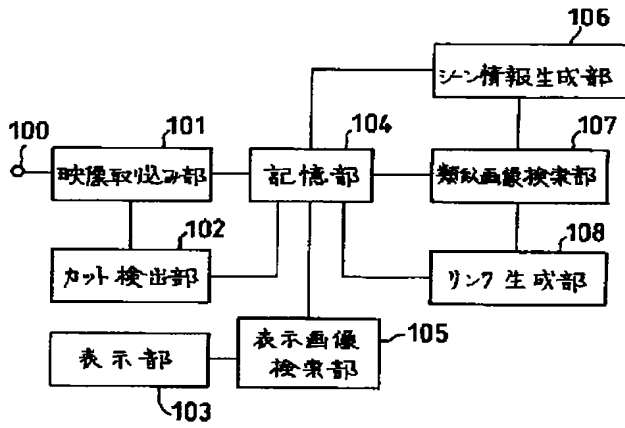
【図11】



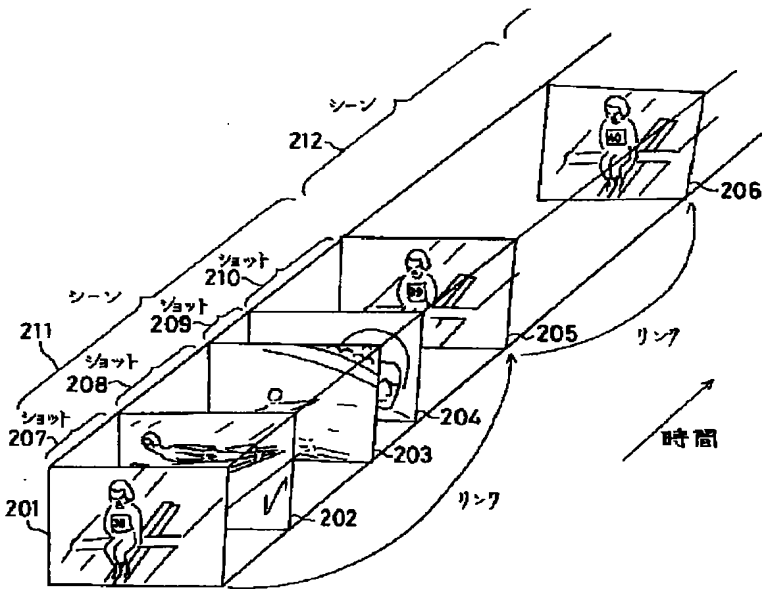
【図12】



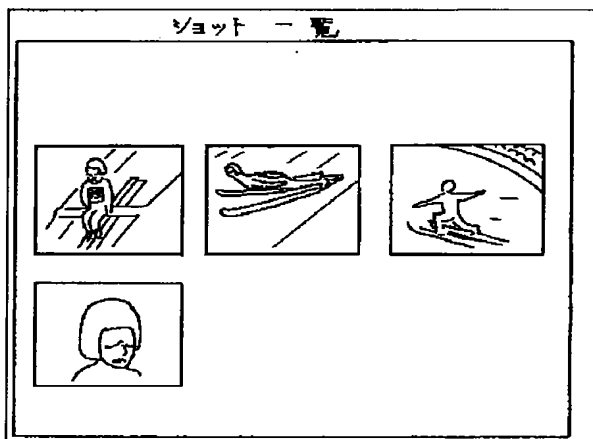
【図13】



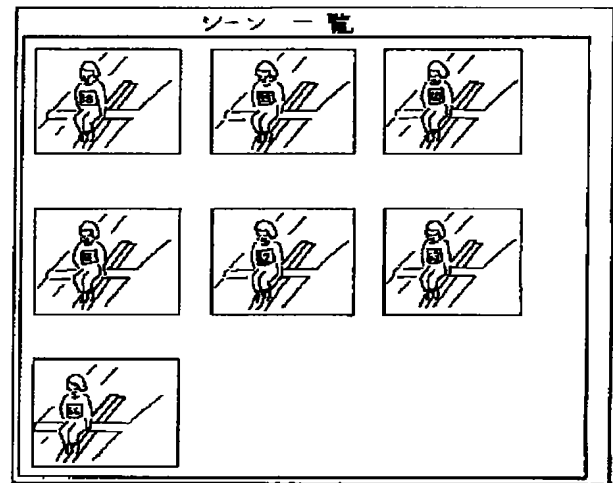
【図14】



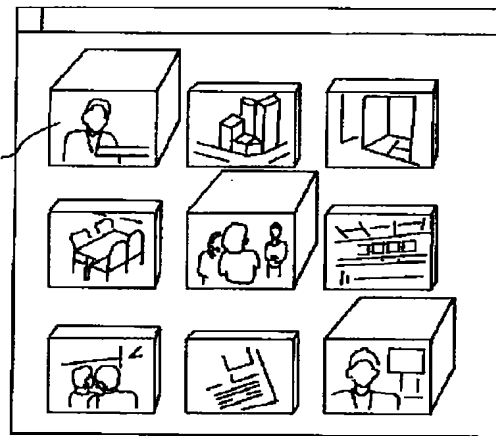
【図16】



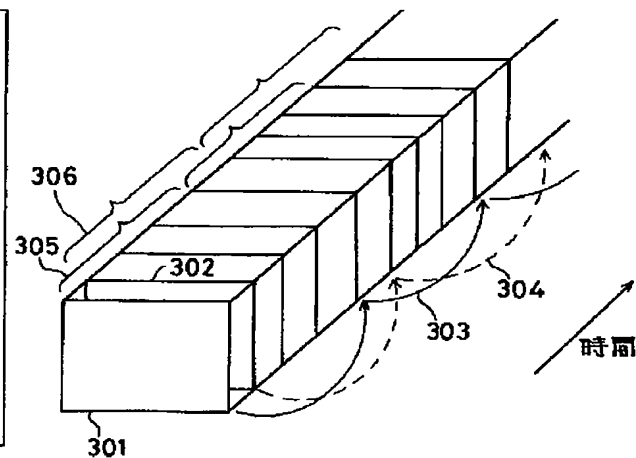
【図15】



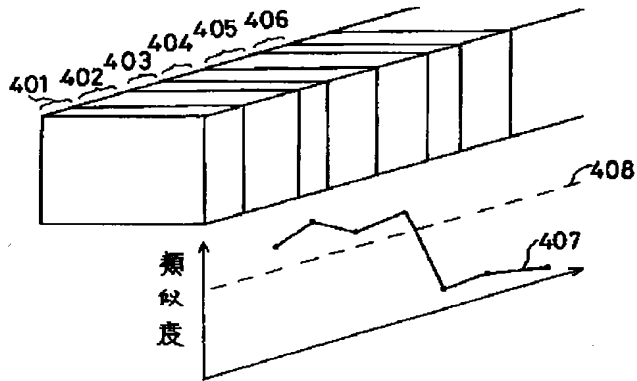
【図23】



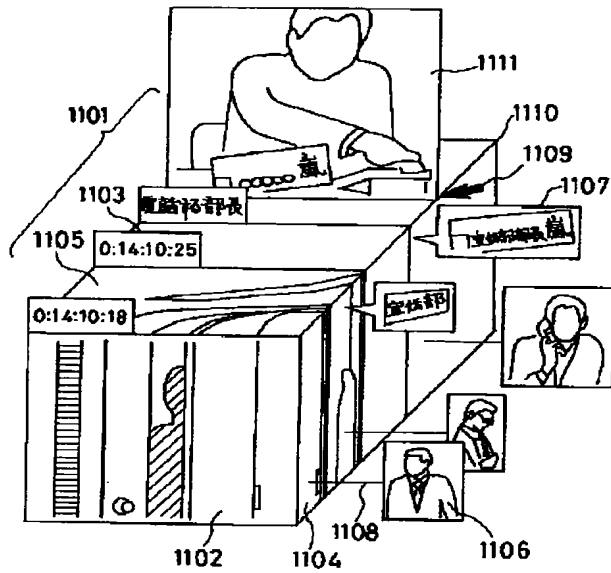
【図17】



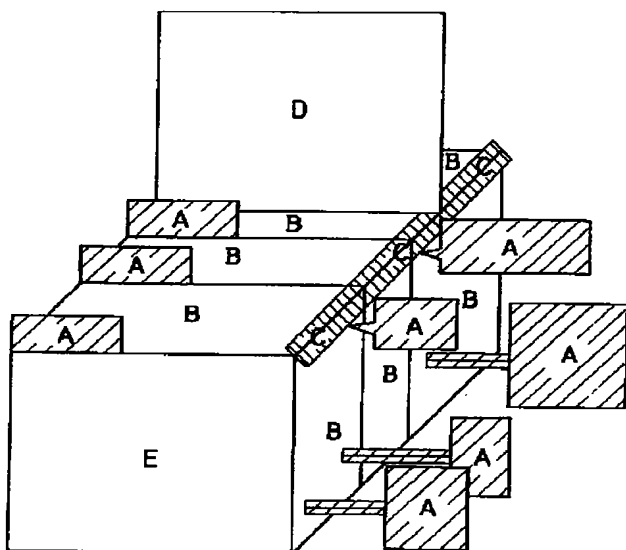
【図 18】



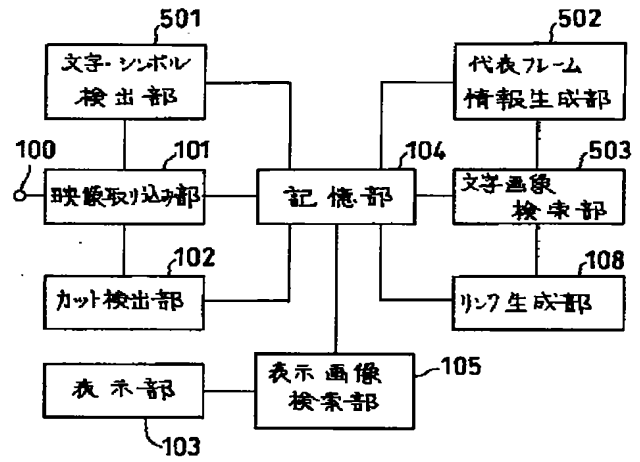
【図 22】



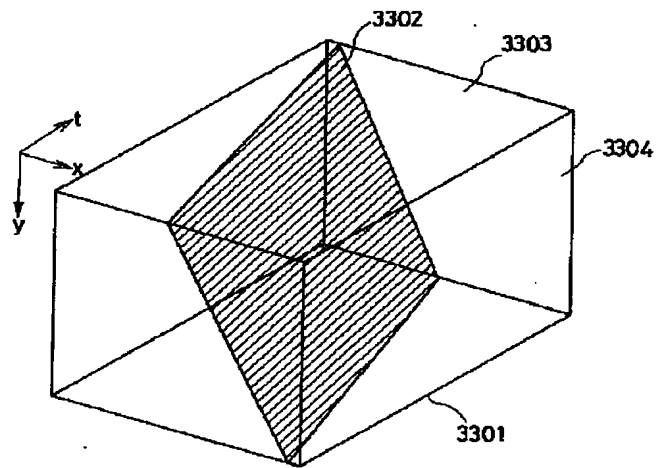
【図 25】



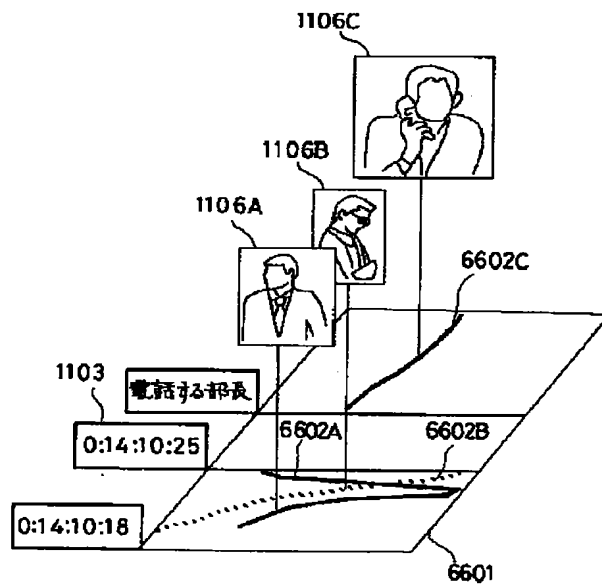
【図 19】



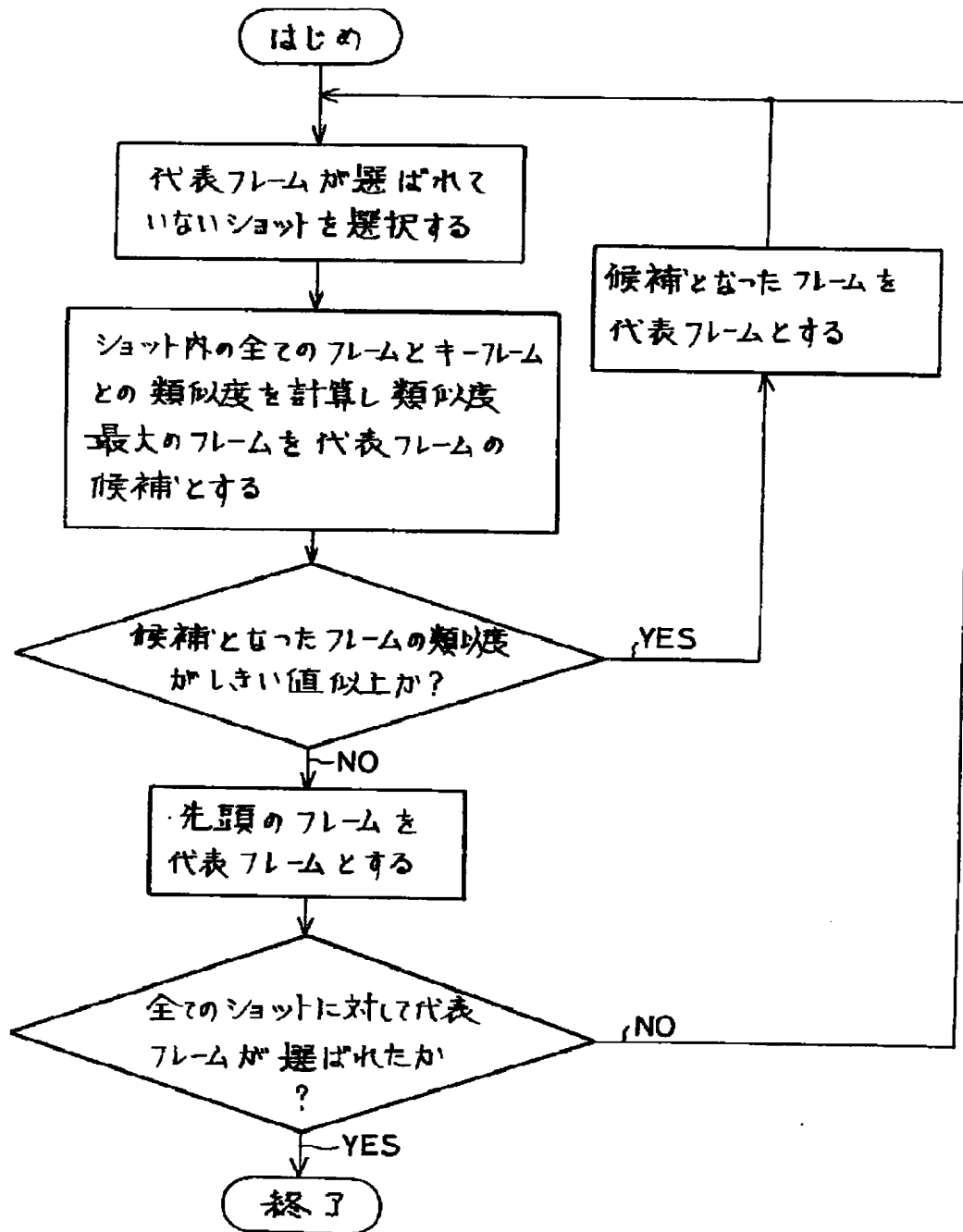
【図 24】



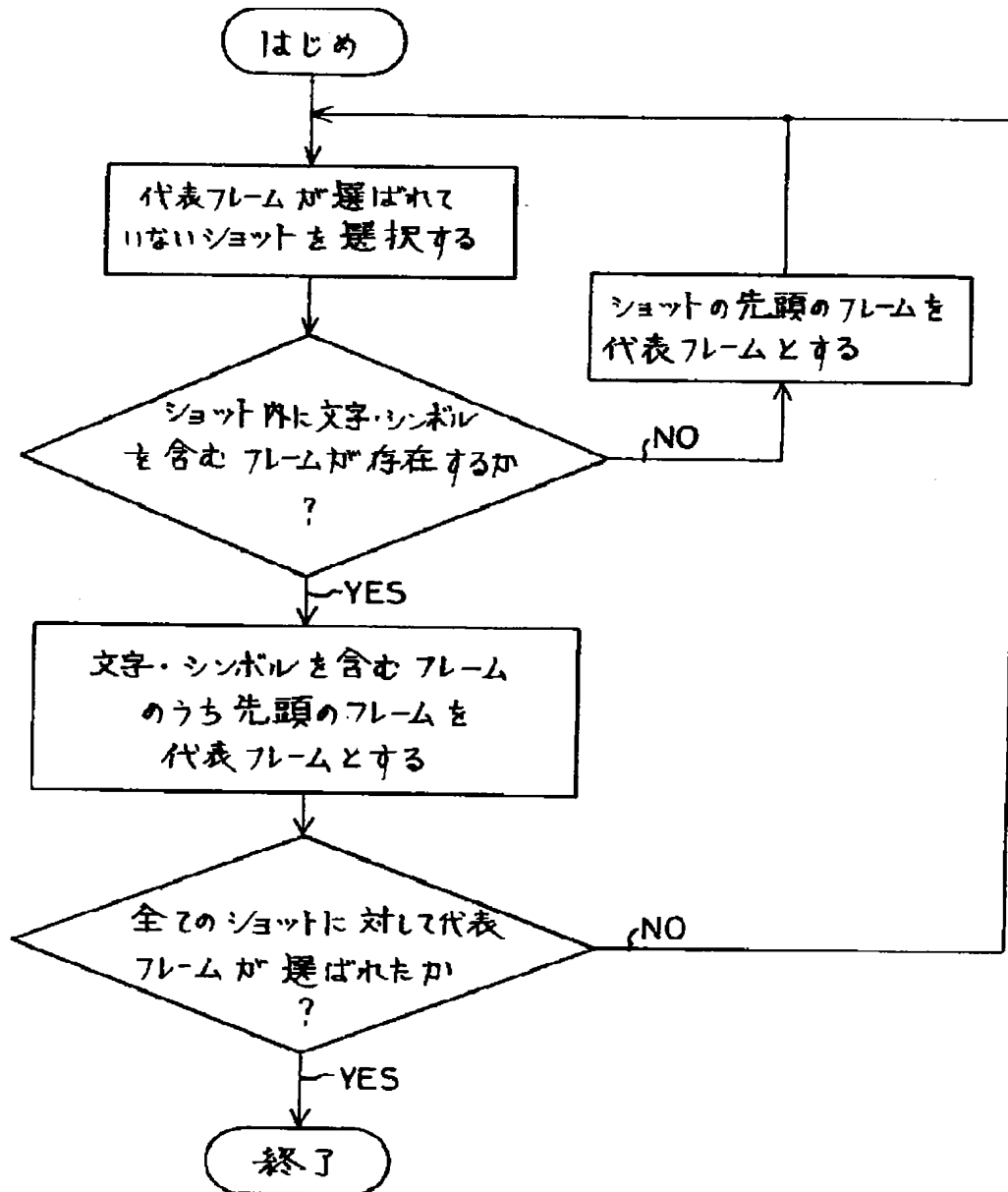
【図 28】



【図20】



【図 21】



【図27】

